Neumann, Irene; Duchhardt, Christoph; Grüßing, Meike; Heinze, Aiso; Knopp, Eva; Ehmke, Timo

Modeling and assessing mathematical competence over the lifespan

Journal for educational research online 5 (2013) 2, S. 80-109

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 5 (2013), No. 2, 80–109

Irene Neumann, Christoph Duchhardt, Meike Grüßing, Aiso Heinze, Eva Knopp & Timo Ehmke

Modeling and assessing mathematical competence over the lifespan

Abstract

Mathematical literacy is regarded as an important prerequisite to mastering problems of everyday life. In the German National Educational Panel Study (NEPS), mathematics has therefore been included as a central domain of competence development over the lifespan. To track the development of mathematical competence in individuals, instruments are needed that provide coherent and consistent measures. The instruments are based on a theoretical framework of mathematical competence over the lifespan. The framework consists of a content-related and a cognitive dimension. The content areas differentiate between four overarching ideas of mathematics. The cognitive component consists of six cognitive processes that are needed to solve mathematics-related problems. Following this structure, the NEPS framework for mathematical competence is compatible with the underlying framework of the PISA studies and with the framework of the German Mathematics Education Standards. The main focus of the manuscript is to accurately describe the NEPS framework of mathematical competence over the lifespan. First, the concept of mathematical competence, on which the NEPS mathematics tests are based, is explained in detail. Then, exemplary items for different age groups illustrate the interplay of content areas and cognitive components. Finally, initial insight into the tests' quality is provided on the basis of pilot studies in Grade 9 of secondary school and in the adult samples.

Juniorprof. Dr. Irene Neumann (corresponding author), Department of Physics and Astronomy, Ruhr-University Bochum, Universitätsstraße 150, 44801, Bochum, Germany
e-mail: neumann@physik.rub.de

Dr. Christoph Duchhardt · Dr. Meike Grüßing · Prof. Dr. Aiso Heinze · Dr. Eva Knopp, Leibniz Institute for Science and Mathematics Education (IPN) at the University of Kiel, Olshausenstraße 62, 24098 Kiel, Germany
e-mail: duchhardt@ipn.uni-kiel.de
gruessing@ipn.uni-kiel.de
heinze@ipn.uni-kiel.de
knopp@ipn.uni-kiel.de

Prof. Dr. Timo Ehmke, Institute of Educational Science, Leuphana University Lüneburg, Scharnhorststr. 1, C1.231, 21335 Lüneburg, Germany
e-mail: tehmke@leuphana.de

Keywords
Mathematical literacy; Mathematical competence; Assessment; Lifespan

Modellierung und Erfassung mathematischer Kompetenz über die Lebensspanne

Zusammenfassung
Mathematical Literacy wird als wichtige Voraussetzung zum Lösen mathematikbezogener Alltagsprobleme angesehen. Daher ist im Projekt „NEPS – Nationales Bildungspanel“ Mathematik als eine zentrale Domäne in der Untersuchung von Kompetenzentwicklung über die Lebensspanne berücksichtigt. Um die Entwicklung mathematischer Kompetenz von Individuen verfolgen zu können, werden Testinstrumente benötigt, die dieses Konstrukt kohärent und konsistent messen. Diese Instrumente basieren auf einer theoretischen Rahmenkonzeption mathematischer Kompetenz über die Lebensspanne. Die Rahmenkonzeption unterscheidet eine inhaltliche und eine kognitive Dimension. In den Inhaltsbereichen werden vier mathematische Leitideen berücksichtigt. Die kognitive Komponente besteht aus sechs Prozessen, die zum Lösen mathematischer Probleme notwendig sind. Mit der Unterscheidung dieser zwei Dimensionen ist die NEPS-Rahmenkonzeption anschlussfähig an die Konzeptionen der PISA Studien und der Bildungsstandards für Mathematik. Im Fokus des Artikels steht die ausführliche Beschreibung der NEPS-Rahmenkonzeption für mathematische Kompetenz über die Lebensspanne. Zunächst beschreibt der Artikel detailliert die Konzeption mathematischer Kompetenz, die den NEPS-Mathematiktests zugrunde liegt. Anhand von Beispielitems für verschiedene Altersgruppen wird das Zusammenspiel von Inhaltsbereichen und kognitiver Komponente veranschaulicht. Schließlich werden auf Grundlage der NEPS-Pilotstudien der Klasse 9 und der Erwachsenen erste Hinweise auf die Qualität der Tests dargestellt.

Schlagworte
Mathematical Literacy; Mathematische Kompetenz; Diagnose; Lebensspanne

1. Introduction

How much paint do I need to decorate my room? How much is the price of a skirt on a 20% off sale? What is the probability of developing side-effects in connection with taking a particular medicine? How does the car rental rate change with respect to the number of days in the lease? Answering these questions involves a mathematical approach; they illustrate the large variety of mathematical problems people encounter in their daily life. Given that education in school aims to prepare students for their future lives, mathematics is viewed as a central component of education: Baumert (2002) includes mathematizing as one of five basic competencies that contribute to general education. In this context, mathematization might be understood as transferring a problem from a real situation into the language of mathematics, solving this mathematical problem with the tools of mathematics, and then evaluating the result with respect to the real-world problem (Baumert, 2002). In this sense, mathematization is equivalent to mathematical modeling as defined in the literature of mathematics education (e.g., Blum & Leiß, 2005; Borromeo Ferri, 2006; Penrose, 1978).

The fact that mathematics is generally regarded as an important educational topic is also reflected by its inclusion in recent international large-scale studies such as the Trends in International Mathematics and Science Study (TIMSS; e.g., Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009) and the Programme for International Student Assessment (PISA; e.g., OECD, 2003). The frameworks for both of these studies – TIMSS and PISA – build on the fact that mathematics plays a central role in people's daily lives. Certainly, this is also reflected in the important part mathematics plays in educational documents such as the Principles and Standards for School Mathematics (NCTM, 2003) and the German Mathematics Education Standards (GMES; KMK, 2004, 2005).

Due to the central role of mathematics in educational contexts, mathematics forms one key competence domain in the National Educational Panel Study (NEPS) as well (Weinert et al., 2011). In contrast to other comparative, large-scale studies such as TIMSS or PISA, NEPS does not only focus on one or a few discrete age cohorts but traces the competence development of individuals over the lifespan. As a consequence, a framework for mathematical competence is required that allows for a coherent and consistent assessment of people's competence from childhood to adulthood. The main goal of this article is to provide a theoretical description of the NEPS framework for mathematical competence, which underpins the NEPS instruments. First, a literature review presents recent approaches to frame mathematical capabilities. On the basis of these approaches, the NEPS framework for mathematical competence is elaborated further. Sample items will illustrate the framework's operationalization in different age cohorts. Finally, results from the ninth graders' and the adults' pilot studies are presented, which provide initial evidence on the quality of NEPS instruments.

2. Approaches to frame mathematical competence

International large-scale studies such as TIMSS (Mullis et al., 2009) and PISA (OECD, 2003) as well as educational documents such as the NCTM standards (NCTM, 2003), the GMES (KMK, 2004, 2005) or the documents provided by the Common Core State Standards Initiative[1] each include an outline of mathematical abilities and skills. These approaches can be classified with respect to their de-

1 http://www.corestandards.org

gree of adherence to the school curriculum, on the one hand, and the literacy concept characterized by its relevance for everyday life on the other hand. Educational documents, such as the NCTM standards, the GMES, and the Common Core State Standards, describe the abilities that students ought to achieve through schooling. By contrast, the PISA framework takes the perspective from the students' future lives, thus focusing on the abilities that students will need to handle problems in their daily lives. In the TIMS studies, the framework of mathematical literacy is based on a "curriculum model" (Mullis et al., 2009, p. 10) and, therefore, is more closely related to the school curriculum. In the following paragraphs, two frameworks for mathematical competence are described in more detail: the PISA framework for mathematical literacy representing the perspective of relevance to everyday life and the German Mathematics Education Standards' framework for mathematical competence representing the school-curriculum perspective.

2.1 The PISA framework for mathematical literacy

The PISA studies aim at assessing how well 15-year-old students are prepared for their future lives – as individuals and as actively participating members of society (OECD, 1999). Accordingly, mathematical literacy as assessed in PISA is defined as

> [...] an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen. (OECD, 2003, p. 24)

This definition reflects that PISA does not merely assess to what extent students have learned the content of school curricula but rather focuses on the students' abilities to mathematically solve problems that are relevant to their daily lives. According to PISA, a problem and its solution is made up of three components: (a) the context of a problem, (b) the content addressed in a problem, and (c) the competencies necessary to solve a problem (OECD, 2003). To account for mathematical literacy as a relevant aspect of everyday life, the contexts chosen for the problems have got to be authentic. This means that problems – even if they include fictitious or hypothetical elements – should relate to real life instead of "being merely a vehicle for the purpose of practicing some mathematics" (OECD, 2003, p. 33). A problem's content is classified by four so-called "overarching ideas" (OECD, 2003, p. 34): *quantity*, *space and shape*, *change and relationships*, and *uncertainty*. Through these ideas a wide range of mathematical topics is addressed without being too specific and losing the close reference to daily life. Finally, eight competencies are differentiated, which constitute cognitive processes needed for mathematical modeling and problem solving: *thinking and reasoning; argumentation; communication; modeling; problem posing and solving; representation; using symbolic, formal and technical language and operations*; and *use of aids* *and tools*. The cognitive demands related to these competencies might vary from problem to problem: Each competence addresses a wide ability range from applying routines to using complex higher order thinking skills (cf. OECD, 2003). Note that the most recent PISA study of 2012 conceptualized mathematical literacy as seven so-called "fundamental mathematical capabilities" (p. 30): *communication; mathematising; representation; reasoning and argument; devising strategies for solving problems; using symbolic, formal and technical language and operations; using mathematical tools* (OECD, 2013).

2.2 Mathematical competence according to the German Mathematics Education Standards

The German Mathematics Education Standards (KMK, 2004, 2005) are based on the PISA framework, yet they adhere more closely to the school curriculum. As Germany is a Federal Republic made up of 16 Federal States (Länder), which each have "legislative as well as administrative competency" with respect to education (Döbert, 2007, p. 300), the GMES have been introduced to harmonize the output of schooling across all Länder. The standards describe competencies that shall provide students with the capability to cope with daily life experiences from a mathematical perspective – such as operating social and cultural processes by means of mathematics or using mathematics to solve problems (KMK, 2004). According to the GMES, two types of competencies contribute to such capabilities: *content-related mathematical competencies* and *general mathematical competencies*. Similar to the PISA framework, the content-related mathematical competencies are classified according to five overarching ideas – that is, *number, measuring, space and shape, functional relationship,* and *data and chance*[2] (KMK, 2004). General mathematical competencies comprise cognitive processes required to approach a task mathematically: *mathematical argumentation*; *mathematical problem solving*; *mathematical modeling*; *use of mathematical representations*; *dealing with symbolic, formal and technical elements of mathematics*; and *mathematical communication*[3] (KMK, 2004). The GMES emphasize that the general mathematical competencies might not be viewed as separate from each other but rather as interconnected with one another. Finally, the GMES differentiate between three cognitive levels on which the general mathematical competencies can be activated: (a) *reproducing,* (b) *establishing connections,* and (c) *generalizing and reflecting* (KMK, 2004). Similar to the differentiated competencies in PISA, the three GMES levels cover a wide ability range. In summary, the GMES framework is very similar to

2 These five ideas are named in the standards for secondary school; those for elementary school are slightly different: numbers and operations; space and shape; patterns and structures; units and measuring; data, frequency and probability (KMK, 2005).

3 These six competencies are named in the standards for secondary school; those for elementary school are similar with the exception of dealing with symbolic, formal and technical elements of mathematics (KMK, 2005).

the PISA framework. The GMES, however, describe the outcomes of mathematics teaching in Germany and, thus, imply consequences for the curriculum, whereas PISA takes a backward perspective from the required abilities and skills of the students' future daily lives.

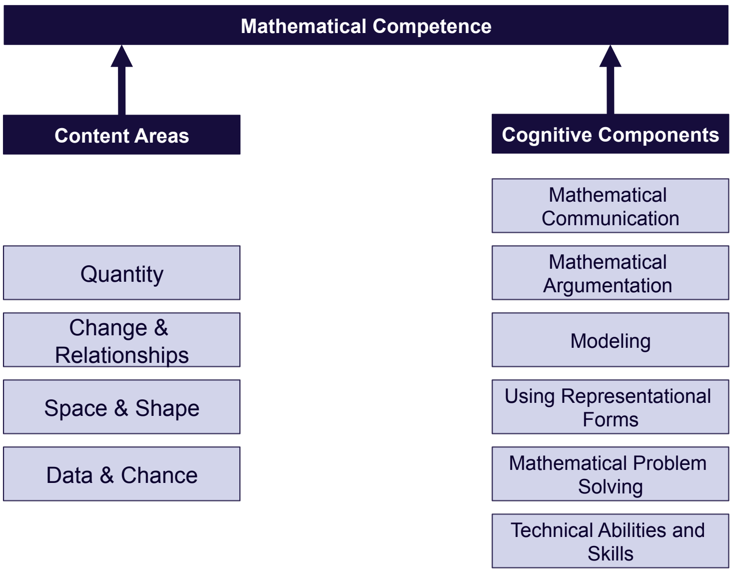
3. The NEPS framework for mathematical competence

With respect to competence development, the NEPS project aims to investigate questions such as the following: How does competence develop over the course of life; how are competence, learning environments, and educational decisions interrelated; how does competence development influence returns to education (cf. Weinert et al., 2011)? Many educational decisions – for example, those referring to schooling – are closely connected to the curriculum and to educational conditions in Germany. The German Mathematics Education Standards provide the current setting for mathematics education in school. Accordingly, the NEPS framework for mathematical competence needs to be compatible with the GMES in order to approach research questions on educational decisions appropriately. Additionally, the NEPS framework for mathematical competence may not only take into account schooling time but must cover mathematical situations in adulthood as well – in particular to approach questions as to what competencies are specifically needed for a successful professional life, and how competencies achieved in school influence further education. As a consequence, the framework also adapts the concept of mathematical literacy as used in the PISA studies (OECD, 2003).

In order to be compatible with the curricular view (as in the GMES) and with the literacy view of mathematical competence (as in PISA), the NEPS framework has been developed in very close connection with both the GMES framework and the PISA framework. This closeness becomes most evident in those age groups that are covered by NEPS, on the one hand, and by GMES or PISA, on the other hand – that is, in the groups of Grade 4 and Grade 9 students (or 15-year-olds). The remaining gaps (early childhood to Grade 4, Grade 4 to Grade 9, Grade 9 to adulthood) have been filled in a coherent manner. In summary, the concept of mathematical competence in NEPS could be described as relevant for future life and, to a limited extend, as curriculum-based as well. The main difference between the NEPS framework for mathematical competence on one side and the GMES and PISA framework on the other side lies in the fact that the GMES and PISA are related to single, selective measures (GMES: Grade 4 and 9; PISA: 15-year-olds), whereas the NEPS framework is required to cover the whole lifespan in a consistent and coherent manner.

Both frameworks of the GMES and PISA are characterized by a differentiation of two components – one related to the content and one related to cognitive processes. On the basis of considerations stated above, the NEPS framework for mathematical competence is structured in the same way (cf. Figure 1): One dimension represents the so-called *content areas*, which cover the field of mathematics, and another dimension represents the *cognitive component* of mathematical competence, which includes cognitive processes that are necessary for solving mathematical problems. To ensure the compatibility between NEPS, on the one hand, and the GMES and PISA, on the other hand, content areas as well as the cognitive component of mathematical competence are conceived as a combination of the respective dimensions from the GMES and PISA. Time (i.e., age group) is viewed as an implicit, third dimension in the NEPS framework. Accordingly, content areas and cognitive component take on age-specific meanings, as described in the following sections[4]. Including this time-related perspective (i.e., age-specific meanings) is the crucial difference between the NEPS framework and the GMES or PISA framework.

Figure 1: NEPS framework for mathematical competence



3.1 Content areas

The discipline of mathematics covers a wide field including geometry, algebra, analysis, probability theory, and so on. Everyday problems involving mathematical aspects often cannot be classified according to this canonical categorization but refer to aspects of several mathematical subareas. The NEPS framework's content areas, therefore, do not follow the canonical classification but refer to four overarching ideas, which are relevant to everyday problems. Accordingly, four content ar-

4 The description of content areas and cognitive component has been taken in large parts from Ehmke et al., 2009.

eas are differentiated: *quantity*, *change and relationships*, *space and shape*, and *data and chance*. Within each content area, a development over time has to be considered. Such development becomes evident in the fact that each content area covers mathematical concepts and procedures that are usually achieved at a particular age. Those mathematical concepts and procedures are typically learned in school and, thus, follow a particular curriculum. However, this does not mean that NEPS test items merely serve to assess how well persons have mastered a specific school curriculum. Rather, the concepts and procedures are embedded in everyday-life contexts that are typical for that particular age group – thus covering the literacy aspect of mathematical competence as well.

In the following, each content area is explained in detail for four age groups: early childhood including kindergarten age, elementary-school age, secondary-school age, and university-students' age. The development described is similar to those found within the NCTM standards' framework (NCTM, 2003). For the adult group, the content areas are presented separately, as we assume that mathematical competence is a conglomerate of what has been learned up to that age. The following description is not based on empirical evidence. Rather, it has been developed in a normative way: It is assumed that the mathematical concepts and processes depicted here are typically achieved by the respective age group.

3.1.1 Quantity

The quantity content area is related to using numbers and quantities in age-specific contexts and situations. Different representations of numbers and relationships among numbers as well as the application of basic mathematical operations are central elements of this content area.

Compared to other content areas, quantity plays the most important role in kindergarten. At this age, a first notion of the concept of numbers has been developed. In kindergarten and early elementary-school age, quantity is split into two subareas: (a) *sets, numbers, and operations*; and (b) *units and measuring*. Building on basic cognitive abilities – such as classifying, seriation, and one-to-one correspondence – the NEPS kindergarten test assesses central mathematical concepts. The subarea sets, numbers, and operations contains concepts such as comparisons of sets ("Are there more red buttons or black buttons?"), representations of numbers ("What numbers relate to this domino tile?"), and counting tasks referring to cardinal aspects ("How many sheep are in this picture?"), as well as ordinal aspects ("Show us the fifth sheep in this row!"). The subarea units and measuring includes concepts such as ordering ("Which of these pencils is the longest?") as well as using measuring tools such as rulers or one's own step length.

Up to Grade 5, children should be able to elaborate their understanding of numbers. At this age, children ought to understand the decimal numeral system; they should be able to compare and order integers and decimal numbers. Moreover, they should understand elementary arithmetic and the corresponding computational strategies and laws, and they ought to be able to compute fluently by means of written algorithms. This content area also includes the relationship between operations (e.g., purposefully using division as inverted multiplication) as well as using estimations to solve authentic problems and checking if such estimations are reasonable.

During lower secondary school (i.e., up to Grade 9), students should develop a meaningful understanding of natural numbers, integers, rational and irrational numbers, as well as the ability to apply this understanding to solving problems associated with them. This ability includes the application of arithmetic laws (computational tricks), algorithms, and controlling strategies such as estimations. Particular problems require students to use calculations of percentages and interests appropriately; other contexts – for example, scientific contexts – require students to correctly apply measures of length, area, or volume and to choose units and quantities (in particular concerning time, mass, money, length, area, volume, and angle).

University students are expected to be familiar with various representations of numbers and should obtain a comprehensive understanding of the properties of numbers, numeral systems, and number systems. For instance, they ought to be able to use matrices for solving systems of linear equations. Moreover, this area includes understanding the concepts of permutation and combination as systematic techniques of counting and to apply these concepts in mathematics-related problems. University students should also be able to make use of integrals for determining a particular content. Depending on the problem, they should be able to decide whether a problem's solution requires a rough estimation, a numerical approximation, or an analytic approach.

3.1.2 Change and relationships

The change-and-relationship area contains problems that require understanding and using (functional) relations between mathematical objects and patterns. Mathematically competent people should be able to analyze quantitative relationships, which are embedded in age-specific problems, and to express these relationships with the means of algebraic symbols.

In kindergarten, this includes identifying and continuing patterns as observed in specific settings, thereby relying on basic abilities such as classifying. In familiar situations, children should also be able to identify basic relationships (e.g., predecessor and successor) and simple proportionalities (e.g., the longer the way, the more steps are needed) or inverse proportionalities (e.g., the more children, the less candy per child). Finally, they should be able to describe qualitative changes in kindergarten age-specific contexts (e.g., different children grow at different speeds).

In Grade 5, change and relationships includes understanding abstract rules of calculation, identifying the rules of geometric and arithmetic patterns, as well as continuing, developing, and changing such patterns. Students should understand variables as an unknown quantity that is represented by a symbol or sign. Moreover, they should show a rudimentary understanding of functional relationships. Competent students should be able to recognize the relationship between the change of one variable and the change of another variable (e.g., the relation between quantity of goods and sum price, proportionality problems, etc.).

In Grade 9, this content area relates to applying functions as a means to describe relationships, to express relationships (verbally; using tables, graphs, or symbols) and to interpret relationships. Students should know characteristic features of functions and should be able to relate functional terms and their graphs. They should be able to solve everyday problems using proportional and inverse proportional relationships as well as linear, quadratic, or exponential functions. In doing so, students should apply algorithms for equation solving.

University students should obtain an elaborate understanding of functions, which allows them to use functions and local changes from an analysis perspective (i.e., determining derivations, roots, maxima and minima, and making use of different representations of functions). They should be able to express problems by means of equations, inequations, and simultaneous equations, and are expected to solve these. Moreover, they should be able to analyze changes embedded in various everyday contexts such as economy, environment, or medicine.

3.1.3 Space and shape

Space and shape relates to any type of planar or spatial configurations, forms, and patterns. This includes analyzing characteristics of geometric forms and objects as well as describing geometric relations. Solving age-specific problems of this content area requires understanding and modeling geometric representations and expressions.

At kindergarten age, holistically identifying geometric forms such as circles, triangles, and squares is central to this content area (e.g., tasks such as “Show me the triangle!”). Moreover, children should be able to analyze forms and to identify their characteristics (e.g., the number of vertices, curved or straight, or magnitudes of angles). This content area also includes congruent mapping, in particular translations (e.g., translations of planar patterns: “Continue this pattern!”). Furthermore, abilities of the space-and-shape content area, which are specific to the kindergarten age, comprise identifying geometric forms and structures in the children’s environment.

In Grade 5, students should obtain a sense of spatial orientation, for example, in order to recognize spatial relationships shown in building plans or perspective drawings, or in order to use nets or edge models. Students should be able to classify geometric bodies and plane figures with respect to their characteristics. Moreover, they should be able to assign technical terms to such objects. Items might require the students to compare two- (or three-) dimensional shapes by dissection or determine their area (or volume) with respect to unit squares (or unit cubes). Students should be able to identify and conduct geometric mappings such as reflections, rotations, and translations; furthermore, they should be able to apply the concept of symmetry.

In Grade 9, this content area includes recognizing, analyzing, and describing plane and spatial geometric structures in the environment as well as mentally operating with line segments, shapes, and bodies. Students should be able to describe and express geometric figures using Cartesian coordinates. They should be able to identify characteristics and relations between geometric objects (e.g., symmetry, congruence, similarity, etc.) and to give reasons for this. Moreover, the students need to know theorems of elementary geometry in order to solve geometric problems in authentic contexts.

University students' mathematical competencies concerning space and shape include the ability to make use of the Cartesian coordinate system and to express geometric mappings of plane objects (e.g., translation, reflection, rotation, scaling) in different ways as, for example, by means of vectors and matrices. They should also understand basic trigonometric concepts. Moreover, students of that age should be familiar with three-dimensional mathematical objects and their relationships (e.g., planes and lines in a three-dimensional space, spheres, etc.).

3.1.4 Data and chance

The content area data and chance focuses on phenomena and situations involving statistical data or chance. Handling data includes collecting information, arranging data, and representing it graphically. Another aspect is to analyze data in problem situations in order to draw conclusions and base predictions on these data. Furthermore, the subarea chance deals with establishing an understanding of the concept of probability and applying it in contexts.

Already in kindergarten, the area of data and chance plays some part. For instance, children should be able to collect objects according to a number of defined criteria (e.g., color of marbles) and to record their number when, again, the ability to classify is asked for. These data might then be organized and interpreted, for example, recording absolute frequencies. An intuitive understanding of chance (e.g., the insight that a certain event is impossible, or else, more likely to occur than another event) should be seen in everyday situations, for example, in dice games.

In Grade 5, children should be able to deal with data more systematically and purposefully than in kindergarten. Competence in this area is indicated by the extent to which children are able to collect data from simple experiments or observations and represent them in tables or figures such as bar charts or line charts. In the subarea "chance" it is required to compare the probabilities of different events in random experiments and to know the basic concepts of "certain", "impossible", or "likely". Children should also be able to assess winning chances in dice games.

In Grade 9, students should be able to plan simple statistical studies, measure data systematically (e.g., distances covered by paper planes with different characteristics), organize data, and represent them graphically (e.g., by histograms or scatter plots). In order to analyze data, students of that age should be able to choose and apply suitable statistical methods (e.g., means or variance). This includes, for example, making conjectures on possible correlations between characteristics of a sample that are based on scatter plots. Students should be able to describe simple random experiments and chance phenomena in daily life mathematically and determine the probabilities of their occurrence.

In higher education, students should have acquired a deeper understanding of statistical surveys. This becomes evident in the knowledge of different research designs (survey, experiment) and the kinds of inference that can be drawn from them. Besides, competent students should be familiar with the basic statistical methods needed to analyze data. They should also be able to judge how adequate an analysis and its conclusions are, considering the methodical design. Another key aspect is to understand the concepts "sample space" and "probability distribution" and to apply them in various contexts.

3.1.5 Content areas in adulthood

The mathematical content of everyday life problems that adults have to master typically does not exceed the content areas described above. Instead, the range of typical problems in adults' everyday life is wide: (a) Adults encounter a variety of subjects from basic arithmetic to multistep-calculations using scientific notations and from simple descriptive statistics to complex statistics; (b) problems cover a wide spectrum from less to more complex situations. Compared to other age groups, the literacy aspect of mathematical competence becomes most obvious in the adult group. Each content area is strongly related to everyday situations, which require adults to understand the corresponding mathematical concepts and procedures. We expect adults to show a wide variance of mathematical competence: Some mathematical concepts and procedures might be more important for particular careers and thus are bound to typical contexts, whereas others are necessary for mastering typical activities of daily life, such as shopping. The following four examples illustrate the variety of contexts and problems relevant in adulthood. Mathematical competence related to the quantity area is needed when adults have to apply their basic mathematical arithmetic skills, for example, to identify the cheapest flight from one city to another or to calculate the 10% discount on a TV set. Concerning the change-and-relationship area, adults should be able to understand how debts may vary depending on the duration of a credit agreement and its monthly rates. Space-and-shape concepts and procedures are needed when adults have to determine how many buckets of paint they will have to buy in order to decorate a room of a particular geometry. Regarding data and chance, adults should be able to estimate their chances of winning the lottery, for example. In a nutshell, mathematical competence in adulthood is characterized by a strong focus on the literacy aspect as well as by a diversity of problems and contents.

3.2 Cognitive component of mathematical competence

The four overarching ideas illustrated above determine mathematical content and are related to mathematical concepts that continue to develop and elaborate over the lifespan. In order to solve problems within these content areas, cognitive processes have to be applied. Such processes constitute the cognitive component of the NEPS framework. Note that a cognitive process is not bound to any particular content area but crosscuts through all of them. The cognitive component's definition borrows from the GMES, which "determine a student's ability based on those competencies that needed to be activated when solving items" (Blum, Drüke-Noe, Hartung, & Köller, 2006, p. 33, own translation). Six cognitive processes are included: *mathematical communication*, *mathematical argumentation*, *modeling*, *using representational forms*, *mathematical problem solving*, and *technical abilities and skills* (see Figure 1). The following definitions of these processes have been adapted from the definitions given by the PISA framework (OECD, 2003) and the GMES framework (KMK, 2004, 2005). Note that the processes are close to the respective competencies detailed in PISA and the GMES framework. However, since the NEPS study is in large parts restricted to closed-ended item types, activity-based components cannot be covered explicitly. For example, *using mathematical tools* (as included in PISA) such as using an electronic calculator is difficult to assess through paper-and-pencil tests as a stand-alone competence. In school, students are typically allowed to use calculators from Grade 8 on. Therefore, from Grade 9 onward (there is no NEPS mathematics assessment in Grade 8) NEPS participants are allowed to use a calculator during test administration. Accordingly, this competence is spread over all other competencies but is not explicitly addressed by single items and not conceptualized as a stand-alone competence. Moreover, we decided not to differentiate between thinking and reasoning, on the one hand, and argumentation, on the other hand (as done in earlier PISA frameworks). In general, mathematical argumentation processes require specific aspects as described in PISA for the competence of thinking and reasoning (e.g., distinguishing between definitions, theorems, examples, etc., or understanding the limit of given mathematical concepts). Very specific paper-and-pencil items are required to differentiate between these two cognitive processes, which goes beyond the idea of a global construct such as mathematical literacy. Accordingly, the GMES do not differentiate between the two competencies, and the most recent PISA framework also conceptualizes reasoning and argumentation as one capability only (OECD, 2013).

3.2.1 Mathematical communication

On the one hand, this cognitive process relates to an understanding of given information (presented verbally, written, or graphically, etc.) in order to extract relevant mathematical content. To do so, an understanding of technical terms of mathematics is necessary among other things. On the other hand, mathematical communication includes communicating mathematical content, such as results or relationships, with other people in a targeted way. For example, mathematical results need to be summarized using appropriate technical language, or mathematical relationships need to be expressed appropriately using drawings or representations.

3.2.2 Mathematical argumentation

This cognitive process means explicitly using mathematical reasoning and includes the central aspects of the PISA competence of thinking and reasoning. Given argumentations (from concrete up to formal argumentations) need to be understood and evaluated. For example, mistakes in argumentation chains have to be identified. Certainly, this process also includes mathematical proofs as well as actively justifying problem solutions. The NEPS framework for mathematical literacy focuses on comprehending and evaluating statements and reasons with respect to particular problems.

3.2.3 Modeling

This cognitive process – in its sense of mathematizing the real world – makes up a central part of mathematical competence. First, a real-world problem needs to be restricted to relevant parameters (so-called situational model). Next, the situational model is transferred into a mathematical model, within which a mathematical solution is elaborated by applying mathematical techniques or problem-solving strategies. Finally, the achieved mathematical results need to be interpreted and validated with respect to the real situation. This modeling circle (cf. Blum & Leiß, 2005; Borromeo Ferri, 2009) is not entirely being considered in the NEPS framework. Here, the focus is rather on the narrow sense of modeling, that is, on the transition from real world to mathematical representation.

3.2.4 Using representational forms

This cognitive process refers to the ability of understanding and expressing mathematical content in different ways. Mathematical content might be given in a mathematical language verbally or in written form, symbolic, or by means of pictorial representation. Moreover, representations based on specific activities are central to this process, in particular with respect to younger age cohorts. With respect to the NEPS framework, this process means extracting information from given mathematical representations (such as tables, graphs, or diagrams), translating mathematical content from one mathematical representational form to another, or relating different mathematical representational forms to each other.

3.2.5 Mathematical problem solving

This cognitive process includes various problem-solving strategies: for example, convergent and divergent thinking, focusing on special cases, generalizing statements, testing different solutions systematically, tracing problems back to known situations. In the NEPS framework, a particular situation is considered a problem if no solving approach becomes obvious to persons of the respective age group.

3.2.6 Technical abilities and skills

These cognitive processes are needed if solving approaches are known and algorithms can be applied. Such abilities and skills are not restricted to simple techniques but rather include available knowledge, internalized calculation algorithms, and applying tools such as calculators.

4. Operationalization of the NEPS framework for mathematical competence

On the basis of the framework described above, test items are assigned to one particular content area that is central to solving the item. By contrast, items might however require several cognitive components of mathematical competence for a successful solution. Increasing mathematical competence corresponds to more complex contents and higher cognitive requirements. Correspondingly, item difficulty will increase if requirements with respect to content and cognitive components become more complex.

The test items reflect the combination of the curriculum view and the literacy view of mathematical competence, which is inherent in the NEPS framework. On the one hand, the items aim at the person's ability to use their mathematical knowledge and skills to meet real-life challenges. This is reflected by the fact that all items are embedded in the context of daily life or situations typical to the special age cohort. On the other hand, the age-specific mathematics tests cover the relevant mathematical concepts and procedures that could typically be achieved within a special age cohort as described above. However, the items were not developed to assess how well students mastered a specific school curriculum. In the following, four sample items are presented to illustrate the framework's operationalization.

These items cover all content areas and cognitive components; additionally, each item refers to a particular age group.

4.1 Sample Item 1

To account for the literacy aspect of mathematical competence in Kindergarten, items are built around instances close to the children's everyday lives. Accordingly, items in Kindergarten are often material-based and, hence, concrete. Sample Item 1 (Figure 2) shows one such item for Kindergarten children from the content area of sets, numbers and operations. The interviewer tells the child that there are four stones inside a covered bowl. Then she puts three additional stones into it and asks how many there are now. The situation itself – playing with stones, board game pieces, or building blocks – is well-known to young children. Solving this task mainly broaches problem solving as, at that age, no obvious computing algorithm for such a situation is typically available yet. Once an algorithm is found, technical skills are needed as well – that is, applying the number-word sequence and counting, possibly involving fingers. In Kindergarten, this item is embedded in a one-to-one interview. With respect to this particular item, the interviewer would use a real bowl and stones. The item might also be used in Grade 1. When carrying out group assessments in Grade 1, the item is then read out to the children, who are asked to circle the correct answer in a booklet.

Figure 2: Sample Item 1

Sample Item 1: **Stones in the Bowl**		
Age cohort(s)	*Content area*	*Cognitive component(s)*
Kindergarten – Grade 1	Quantity (sets, numbers, & operations)	Mathematical problem solving, technical abilities and skills
There are four stones in this bowl. Now I am adding three more stones. [The bowl is covered so that the child cannot look inside.] Can you tell me how many stones are in this bowl now?		

4.2 Sample Item 2

Sample Item 2 (Figure 3) has been developed for young secondary-school children. The central theme is the – numerical – relation between circumference, width, and breadth of a rectangular area, which is part of the content area space and shape. Similar situations are usually well-known at that age – for example, from measuring playing fields or handicraft work such as constructing kites. The key cognitive requisite for tackling this item is problem solving: The children need to trace back this contextualized situation to the rather abstract visualization of a rectangle, which they have probably learned in school. They further need to relate the information given to this visualization (40 m = perimeter of the rectangle, 8 m = width of the rectangle). Then, a relation between the elements is to be established. Finally, the unknown element (length) must be computed, which involves technical abilities and skills. Older children who are already familiar with the geometrical properties of a rectangle might need only technical abilities and skills, because they can apply an algorithm to this problem.

Figure 3: Sample Item 2

Sample Item 2: **The Fence**		
Age cohort(s)	*Content area*	*Cognitive component(s)*
Grade 5 – Grade 7	Space & shape	Problem solving, technical abilities and skills

Mr. Brown owns a rectangular piece of land and wants to fence it in. He has already made some calculations and hence bought a 40 m fence. The piece of land has a width of 8 m. How long is the land?

☐	5 m
☐	8 m
☐	12 m
☐	16 m

4.3 Sample Item 3

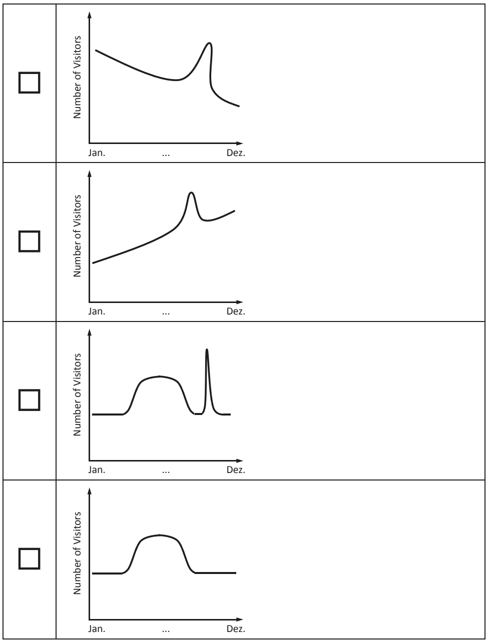
Sample Item 3 (Figure 4) refers to upper secondary school and university students. The item addresses the relationship between time and number of visitors. Therefore, it belongs to the content area of change and relationships. In daily life, change is sometimes expressed by formulas, sometimes by graphs, and other times by a verbal description. Translating between these different representations is a relevant competence for upper secondary students as well as for university students and adults. In this sample item, students have to identify a graph representing a given description of the particular relationship. To do so, they need to understand certain properties of graphs, such as slope, minima, and maxima with respect to different (verbal and graphical) representational forms.

Figure 4: Sample Item 3

***Sample Item 3:* At the Zoo**

Age cohort(s)	*Content area*	*Cognitive component(s)*
Grade 9 – University students	Change & relationships	Using representational forms, communication

In the summertime, the Atown-Zoo is visited by more people than in winter. However this year, a brown bear was born in October, which attracted a lot of visitors due to its cuteness. Which of the following graphs correctly displays the number of visitors at the Atown-Zoo?



4.4 Sample Item 4

Sample Item 4 (Figure 5) has been developed for adults and addresses the content area of data and chance. Empirical data are presented by a contingency table; this is a very common way of representing the relation between two categorical variables in, for example, newspapers. Hence, the item is highly relevant for adults. Adults are asked to evaluate statements on the given empirical data. Accordingly, the main cognitive component that is needed here is argumentation: Are the statements mathematically and logically valid? Do they relate to the data in the table? The argumentation includes aspects of communication as well as technical abilities and skills, because basic calculations have to be made to evaluate the statements with respect to their mathematical validity.

Figure 5: Sample Item 4

Sample Item 4: **Side Effects**

Age cohort(s)	*Content area*	*Cognitive component(s)*
University students – Adults	Data & chance	Mathematical argumentation, communication, technical abilities and skills

A pharmaceutical company has developed a new medicine against headache. A study revealed two major side effects—itching and sickness. The following chart displays the number of participants who were involved in the study.

		Sickness	
		Yes	No
Itching	Yes	50	70
	No	40	100

Are the following statements about the study's result correct?

	yes	no
Half of the participants showed at least one side effect, because 50 is half of 100.	☐	☐
Sickness occurred less than itching, because 50 + 40 is less than 50 + 70.	☐	☐
About 53% of the participants showed at least one side effect, because (50 + 40 + 70)/3 ≈ 53%.	☐	☐
More than half of the participants showing sickness also showed itching, because 50:90 > 50%.	☐	☐

5. Test development

Test development is an iterative process, including expert ratings, pilot studies, and revisions, whose aim is to establish an assessment instrument with good psychometric properties that captures the intended construct (cf. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2004; Hambleton & Zenisky, 2003). In the NEPS project, instruments are embedded in a multicohort sequence design (Weinert et al., 2011); they are employed to track the competence development of individuals. The main challenge of developing such instruments lies in the fact that (a) tests shall provide measures *specifically* for each investigated age group, and (b) tests need to provide a *coherent* measure over the lifespan in order to determine individual competence development. In particular, test instruments developed for adjacent age groups should be directly comparable with respect to the measured construct (cf. the article of Pohl and Carstensen, 2013 in this issue).

In line with the multicohort sequence design, mathematics test items have been developed for several age groups simultaneously. In doing so, large item pools have been designed for each age group. Items are judged and revised by experts; then, field studies are conducted administering about 60 items in each age group. The field studies serve to gain a first insight into the items' difficulty, quality, and appropriateness for the examined age group. On the basis of the field studies' results, about 40 items are then selected to be included in the NEPS pilot studies. Items are selected on the basis of their psychometric quality. Moreover, items are chosen in such a way that content areas are widely balanced. The final instruments still cover cognitive components appropriately; due to limited testing time, however, it has not been possible yet to include every combination of content area and cognitive component. Results from the NEPS pilot studies are used for another item selection, so that mathematics instruments of the NEPS main studies finally contain about 20 items covering all content areas and cognitive components. Similar to the NEPS main study, the NEPS pilot studies also typically follow a multicohort sequence design. This means that – for a limited sample size – initial longitudinal analyses might be conducted (cf. the article of Pohl and Carstensen, 2013 in this issue).

Because the development of mathematical competence in the NEPS framework is considered to be continuous over the lifespan, mathematics instruments need to be coherent as well. Accordingly, possible links between adjacent age groups have already been taken into account during the process of item development. Links are represented by test items that can be used in more than one age group – certainly, item difficulty might not be the same in both age groups. Reflecting the framework over the lifespan, NEPS mathematics tests thus contain items specific to each age group as well as a small amount of items used in different age groups – so-called linking items.

In the following, we focus by way of example on the pilot studies that have been conducted in a Grade 9 and in an adult sample. With respect to the process of item development described above, we will provide initial evidence to approach the following question: To what extent are the NEPS mathematics tests appropriate for measuring mathematical competence (a) in Grade 9 and (b) in the adult sample? This paper does not investigate to what extent the instruments are appropriate for linking different age groups, yet we provide only first insights in the used linking items. This is because we use only data from two pilot studies that would reveal only tentative results. Moreover, NEPS includes so-called *linking studies* that are designed and conducted to investigate this specific question and that will provide more robust evidence (cf. the article of Pohl and Carstensen, 2013 in this issue).

6. Grade 9 and adults' pilot study: Design and sample

To provide initial evidence on the quality of mathematics instruments, two particular age groups have been chosen – Grade 9 students and adults. The Grade 9 and the adult cohort can be viewed as subsequent age cohorts, because in Germany students may start their vocational training after finishing Grade 9. As detailed above, the NEPS framework of mathematical competence has been designed to cover the whole lifespan and is tailored to each specific age group. The adults' test is characterized by a strong focus on the literacy view of mathematical competence; that is, the items are addressing typical everyday problems that the participants have probably encountered before or still encounter in real life. The Grade 9 test carries an additional curriculum perspective that is represented by single items addressing rather academic, abstract problems. Nevertheless, both tests are designed to measure the same construct, as the above framework assumes that mathematical literacy were coherent over the different age cohorts and, in particular, invariant from one to the following age cohort. In order to illustrate this invariance, we have chosen those two age groups for this paper.

For the NEPS Grade 9 and adults' pilot studies 72 items have been used (see Table 1). 11 items were used in both the Grade 9 sample and in the adult sample. An investigation of the linking items provides a first indicator concerning the consistency and coherence between both age groups. In each of the three subsets – Grade 9 only, adults only, and linking Grade 9 and adults – the items covered all content areas. Table 2 illustrates that most of the developed items exhibited a simple multiple-choice format, as shown by way of example in Figure 3 and Figure 4. The other items were either complex multiple-choice items – as shown, for example, in Figure 5 – or required a short answer.

Table 1: Overview of content areas addressed in the Grade 9 and adults studies

Content area	Grade 9 only	Grade 9 & adults	Adults only
Quantity	8	3	6
Space & shape	6	4	6
Change & relationships	10	3	10
Data & chance	7	1	8
	Σ 31	Σ 11	Σ 30

Table 2: Overview of item types used in the Grade 9 and adults studies

Item type	Grade 9 only	Grade 9 & adults	Adults only
Multiple choice	20	9	24
Complex multiple choice	6	1	1
Short answer	5	1	5
	Σ 31	Σ 11	Σ 30

6.1 Grade 9

The Grade 9 pilot study was conducted in winter 2009/10. Test items were compiled into 16 booklets and administered in a classroom setting. The booklets contained a mathematics test and a science test requiring a processing time of 60 minutes each. The mathematics test was administered to one half of the students first, followed by the science test – and vice versa to the other half of students. There was a 15-minutes break between the two tests. The mathematics booklets were divided into two parts with a processing time of 30 minutes each. Inside the booklets, items were rotated to account for position effects; eight different rotations were used. All items were administered to every student. The sample included N = 181 Grade 9 students (92 female, 87 male, 2 nonresponse) from four Federal States in Germany (i.e., Bavaria, Hamburg, North-Rhine Westphalia, and Thuringia). The median reported age of the sample was 14;1 years (min = 11;4 years/max = 18;10 years).

6.2 Adults

The adults' pilot study was conducted in winter 2009/10 as well. The tests were administered in a one-to-one interview setting, mostly at the participants' homes. Test items were compiled into paper-and-pencil booklets using eight rotations to account for position effects. After a standardized introduction, participants took the mathematics test. The interviewers were instructed not to interact with the participants during testing time and to inhibit external distraction as much as possible. The booklets were divided into two parts with a processing time of 30 minutes each. All items were administered to every participant. The sample included N = 461 adults from all over Germany (254 female, 205 male, 2 nonresponse). The median age of the sample was 43;3 years (min. = 19;9 years/max. = 68;7 years). Participants were chosen by taking into account their age and their highest educational degree. Three age groups and three categories of educational degrees were differentiated, resulting in nine (3x3) combinations of age and educational degree (cf. Table 3). Adults were engaged for participation by the interviewer. During this recruiting process, interviewers were asked to cover the nine combinations as well as possible (cf. infas, 2012).

Table 3: Realized adult sampling

Year of birth	Educational degree low	medium	high	Sum
1975–1986	33	57	55	145
1960–1974	46	51	57	154
1959 and older	51	52	59	162
Sum	130	160	171	461

Note. Educational degrees were categorized into three groups: low = Haupt-/Volksschulabschluss, 8th grade of Polytechnische Oberschule, Sonder-/Förderschulabschluss, ohne Abschluss; medium = Mittlere Reife; i.e., Real-/Wirtschaftsschulabschluss, Fachschul-/Fachoberschulreife, 10th grade of Polytechnische Oberschule; high = allgemeine/fachgebundene Hochschulreife, i.e., Abitur/12th grade of Erweiterte Oberschule, and Fachhochschulreife/Abschluss Fachoberschule. Information presented in the table are taken from infas (2012).

7. Grade 9 and adults' pilot study: Methods

All analyses were conducted according to the current common NEPS guidelines (cf. the article of Pohl and Carstensen, 2013 in this issue). Gathered data were analyzed using the unidimensional Rasch model (e.g., Wilson, 2005). All data were scored before the Rasch model was fitted to the data: Simple multiple-choice and open items were coded as correctly or incorrectly; complex multiple-choice items were coded as correctly when all statements were answered correctly. Missing data were treated as not administered, as evidence from simulation studies suggests that this approach is more robust than missing data coded as incorrect responses (Pohl & Carstensen, 2012; Rose, von Davier, & Wu, 2010). Analyses were conducted using the program *ConQuest* (Wu, Adams, & Haldane, 2007). Parameters were estimated using the marginal maximum likelihood method (MML); person ability estimates were calculated using weighted likelihood estimates (WLE). Rasch modeling was used to explore item difficulty, the matching of item difficulties and person abilities, item fit, and reliability. *ConQuest* provides so-called weighted mean square statistics (WMNSQ) indicating an item's fit to the model. Moreover, correlation between a person's score on an item and his or her sum score (so-called discrimination index) indicates an item's power to distinguish between persons. To identify items with good psychometric quality, the following rules of thumb were applied: $0.85 < \text{WMNSQ} < 1.15$ and discrimination $> .3$ (cf. NEPS Technical Report by Pohl & Carstensen, 2012).

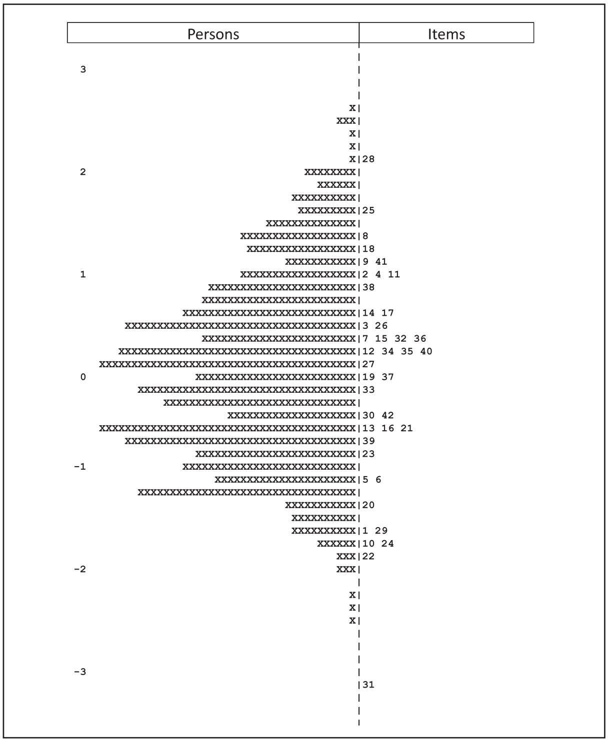
8. Grade 9 and adults' pilot study: Results

To approach our research question, data from Grade 9 and from the adult sample were analyzed separately. The analysis of Grade 9 data included the specific items of Grade 9 and the link items ($n_{\text{items},9} = 42$); likewise, adult data included the specific adults and link items as well ($n_{\text{items,adult}} = 41$).

8.1 Grade 9

Observed Expected A Posteriori/Plausible Value (EAP/PV) reliability was .87, WLE reliability was .86 indicating a highly reliable measurement of students' mathematical competence. Weighted mean square index of the items was between 0.85 and 1.37; discrimination of the items was between .03 and .59. Only three items were misfitting (WMNSQ > 1.15); seven items showed a low discrimination (< .3). The *p* values (item difficulty) of the items ranged between .15 and .94. Observed variance was 0.91. Figure 6 shows the distribution of item difficulties and students' abilities as observed in the Grade 9 sample. The item pool used covered the students' ability range satisfyingly.

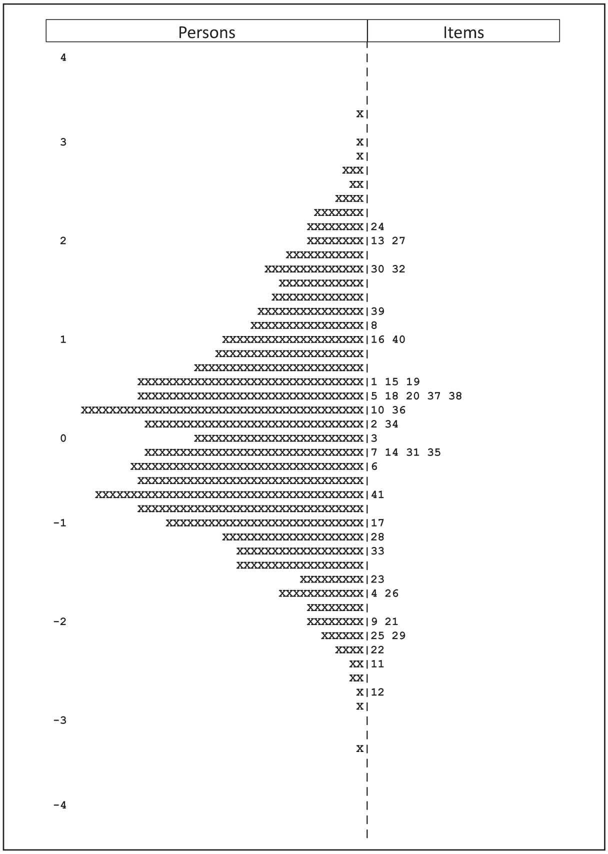
Figure 6: Wright map of Grade 9 data. 'X' represents 0.3 cases; numbers indicate items



8.2 Adults

With respect to the adult sample, observed EAP/PV reliability was .87, WLE reliability was .86. Accordingly, the instrument measured adults' mathematical competence consistently. Exploring item quality revealed that WMNSQ indices of the items were between 0.80 and 1.32; discrimination values were between .06 and .63. Four of 41 items showed unsatisfying model fit (WMNSQ > 1.15); the same items were those exhibiting low discrimination (< .3). The *p* values (item difficulty) of the items ranged between .17 and .91. The observed variance was 1.25. Figure 7 displays the distribution of item difficulties and persons' abilities as observed in the adult sample. The item set used covered the ability range appropriately.

Figure 7: Wright map of the adults data. 'X' represents 0.7 cases, numbers indicate items



8.3 Comparing Grade 9 and adults: Initial insights

The adults' abilities cover a range of about 7 logits. Compared to the Grade 9 sample the adult's range and observed variance is larger; this finding might indicate that the adult sample is a less homogeneous sample. However, the varying range might also have emerged as a result of the instruments having different discrimination. To rule out the latter inference, further evidence from the linking studies is needed.

Figure 8 presents the distribution of *p* values for each content area. For every content area, the two age cohorts are displayed separately. Again, this figure shows that observed variance is larger for the adult group than for Grade 9 students. Quantity items seem to be easiest for both samples. Moreover, data-and-chance problems appear to be easier for Grade 9 students. This finding might possibly come as a result of data-and-chance problems not occurring very often in daily life. Again, these results are rather tentative and should be corroborated through further studies.

A special feature of the reported pilot studies is a set of 11 items, which has been included in both the Grade 9 and in the adults' study. Figure 8 and Table 4 display the *p* values of these linking items for both samples. On each item the adults outperformed the Grade 9 students. This might initially indicate that, in general, the adults show a higher mathematical competence than the Grade 9 students. This tentative inference and more detailed conclusions are expected from the NEPS linking studies.

Figure 8: Distribution of *p* values within the four content areas. Black circles refer to the adult group, blue ones to Grade 9 students. Red dots indicate the linking items. C&R = Change and relationship; D&C = Data and chance; QU = Quantity; S&S = Space and shape

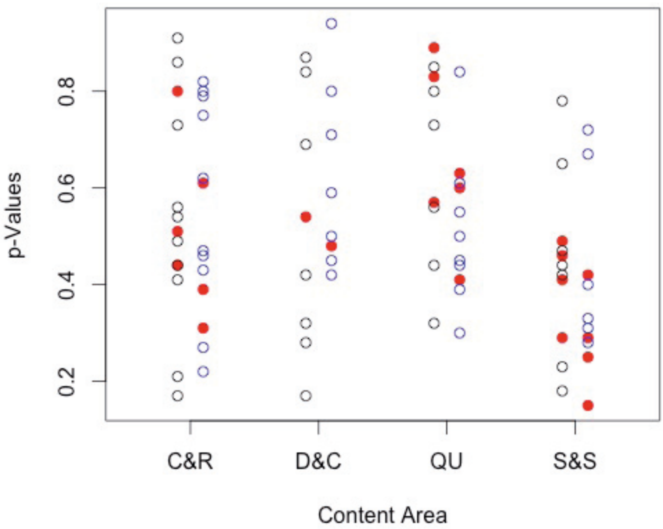


Table 4: Percentage of correct responses to link items

		p values	
Item no.	Content area	Adults	Grade 9
1	Space & shape	.41	.25
2	Space & shape	.49	.29
3	Change & relationships	.51	.31
4	Change & relationships	.80	.61
5	Change & relationships	.44	.39
6	Quantity	.57	.41
7	Data & chance	.54	.48
8	Space & shape	.29	.15
9	Quantity	.83	.60
10	Space & shape	.46	.42
11	Quantity	.89	.63

8.4 Item selection

The above analyses from the pilot study data were used to select items for the NEPS main studies. During item selection, the items' WMNSQ and discrimination as well as their difficulty were taken into account. High-quality items that showed a WMNSQ < 1.25 and discrimination < .30 were chosen for the main study instruments. Difficulties of the selected items were moderate between 0.2 and 0.8 logits. Additionally, wrong answers should show a negative point biserial correlation. Not only were these psychometric criteria used for item selection but the need of covering the theoretical framework also played an important part. That is, items were selected so that they would cover the four content areas and the six cognitive processes evenly, if at all possible. This should ensure the content validity of the NEPS main study instruments. As a consequence, 22 items were included in the Grade 9 test and 22 items in the adults' test for the main studies. In both instruments, eight items were used as linking items connecting the Grade 9 and the adults' cohort. The reduction of linking items was mainly due to reasons of content validity. For all selected items, the distractors were carefully examined. Distractors that had been chosen by only a rather limited number of participants (i.e., unattractive distractors) were subsequently reworded to increase their attractiveness.

9. Discussion

In the NEPS project, mathematical literacy shall be tracked over the lifespan. To develop appropriate assessment instruments, a framework is needed that describes mathematical competence over the lifespan coherently. To account for the curriculum and the educational setting in Germany, the German Mathematics Education

Standards' framework was adapted; to account for compatibility with real life in adulthood, the PISA framework of mathematical literacy was adapted. Merging both frameworks, the resulting NEPS framework for mathematical competence over the lifespan now consists of two dimensions: (a) Content areas describe four overarching ideas that occur in mathematical problems; (b) the cognitive component includes six cognitive processes that are needed to solve mathematical problems. Content areas as well as the cognitive component successively change over the lifespan because they take into consideration the age-related context in which mathematics is used and/or learned. This framework serves as a basis for test item development. Each test item is assigned to one content area and requires one or more cognitive components. On the basis of this framework, NEPS mathematics assessment instruments are being developed for each age group specifically, yet they include a particular overlap between the age groups.

Item development deploys pilot studies in order to determine item quality and appropriateness with respect to a particular age group. On the basis of the Grade 9 and the adults' pilot studies we have illustrated that the developed test items showed satisfying item quality in both age groups. Moreover, the items covered the ability range quite well and provided a reliable measure for mathematical competence. In summary, these results are a first, tentative indicator that, based on the same theoretical framework, the NEPS tests do provide appropriate measures of mathematical competence in different age groups. However, the question of whether the NEPS tests do in fact provide measures of the same construct across the whole lifespan requires further empirical evidence. Within the NEPS project such evidence will be gleaned from conducting specific linking studies (cf. the article of Pohl and Carstensen, 2013 in this issue): Linking studies are conducted to investigate to what extent the tests from different age groups measure the same construct (consistency). Additionally, the longitudinal data will be used to explore how the construct of mathematical competence changes over the lifespan (coherence). The fact that all tests have emerged from the same framework of mathematical literacy is a first step toward such a coherent and consistent measurement of mathematical competence over the lifespan.

Finally, further studies might reveal additional insights into the instruments' validity. With respect to particular age groups, the NEPS mathematics tests could be compared with other standardized instruments (e.g., PISA mathematics tests in Grade 9). Additionally, supplementary studies might reveal an insight into the framework's structure. That is, the framework's dimensions could be investigated with respect to the number of constructs they represent. The cognitive component of mathematical competence differentiates six processes that are needed to conceptualize mathematics problems. However, the NEPS framework assumes that conceptualizing a problem typically requires more than one cognitive component. To analyze the data with regard to whether the cognitive component represents one or more constructs would therefore not be the reasonable thing to do. By contrast, one item in the NEPS mathematics tests addresses only one content area, which gives sense to investigating multidimensionality in principle. However, the NEPS project allows only a limited processing time for assessing mathematical competence, resulting in only a small number of items per content area. Only five to six items per content area can be included in a NEPS mathematics instrument. From a content validity perspective it is highly questionable whether this small amount of items would sufficiently cover the mathematics knowledge taught in several school years. Accordingly, measuring mathematical competence specifically with respect to each content area is not advisable. However, please note that the recently published NEPS Technical Report on the Grade 9 main study reports on multidimensionality analyses indicating very high correlations between the four content areas (Duchhardt & Gerdes, 2013). On the basis of the rationale explained above, this result still calls for corroboration through supplementary studies employing yet a larger number of items.

Acknowledgments

We would like to thank the reviewers for their insightful comments that have improved the quality of our manuscript. Moreover, we would like to thank Dr. Natalie Boonyaprasop for thorough proofreading on English language issues. The studies presented in this paper are part of the National Educational Panel Study and have been funded by the German Federal Ministry of Education and Research. The Leibniz Institute of Science and Mathematics Education supported this work. We thank the infas Institute for Applied Social Sciences for providing information on the adult sample.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2004). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Baumert, J. (2002). Deutschland im internationalen Bildungsvergleich. In N. Killius, J. Kluge, & L. Reisch (Eds.), *Die Zukunft der Bildung* (pp. 100–150). Frankfurt a. M., Germany: Suhrkamp.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (Eds.). (2006). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen*. Berlin, Germany: Cornelsen Scriptor.

Blum, W., & Leiß, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *mathematik lehren, 128*, 18–21.

Borromeo Ferri, R. (2006). Theoretical and empirical differentiations of phases in the modelling process. *Zentralblatt der Mathematikdidaktik, 38*(2), 86–95.

Döbert, H. (2007). Germany. In W. Hörner, H. Döbert, B. von Kopp, & W. Mitter (Eds.), *The education systems of Europe* (pp. 299–325). Dordrecht, Germany: Springer.

Duchhardt, C., & Gerdes, A. (2013). *NEPS Technical Report for Mathematics – Scaling results of Starting Cohort 4 in ninth grade* (NEPS Working Paper No. 22).

Bamberg: University of Bamberg, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/Working%20Papers/WP_XXII.pdf

Ehmke, T., Duchhardt, C., Geiser, H., Grüßing, M., Heinze, A., & Marschick, F. (2009). Kompetenzentwicklung über die Lebensspanne: Erhebung von mathematischer Kompetenz im Nationalen Bildungspanel. In A. Heinze & M. Grüßing (Eds.). *Mathematiklernen vom Kindergarten bis zum Studium: Kontinuität und Kohärenz als Herausforderung für den Mathematikunterricht* (pp. 313–327). Münster, Germany: Waxmann.

Hambleton, R. K., & Zenisky, A. (2003). Advances in criterion-referenced testing methods and practices. In C. R. Reynolds & R. W. Kamphaus (Eds.), *The handbook of psychological and educational assessment* (2nd ed., pp. 377–404). New York, NY: The Guilford Press.

infas – Institut für angewandte Sozialforschung. (2012). *Entwicklungsstudie Etappe 8 Erwachsene, infas_Dokumentation_B60_B61_B62_411362*. Bonn, Germany.

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (Jahrgangsstufe 10)*. Neuwied, Germany: Luchterhand.

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005). *Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4)*. Neuwied, Germany: Luchterhand.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011. Assessment frameworks*. Amsterdam, Netherlands: IEA.

NCTM – National Council of Teachers of Mathematics. (2003). *Principles and standards for school mathematics* (3rd ed.). Reston, VA: National Council of Teachers of Mathematics.

OECD – Organisation for Economic Co-Operation and Development. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris, France: OECD.

OECD – Organisation for Economic Co-Operation and Development. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem-solving knowledge and skills*. Paris, France: OECD.

OECD – Organisation for Economic Co-Operation and Development. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264190511-en.

Penrose, O. (1978). How can we teach mathematical modelling? *Journal of Mathematical Modelling for Teachers, 1*(2), 31–42.

Pohl, S., & Carstensen, C. H. (2012). *NEPS technical report – Scaling the data of the competence tests* (NEPS Working Paper No. 14). Bamberg: University of Bamberg, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/Working%20Papers/WP_XIV.pdf

Pohl, S., & Carstensen, C. H. (2013). Scaling of competence tests in the National Educational Panel Study – Many questions, some answers, and further challenges. *Journal for Educational Research Online, 5*(2), 189–216.

Rose, N., von Davier, M., & Wu, X. (2010). *Modeling nonignorable missing data with item response theory (IRT)* (ETS Research Report ETS-RR-10-11). Retrieved from http://www.ets.org/research/policy_research_reports/publications/report/2010/ibyb

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. H. (2011). Development of competencies across the lifespan. *Zeitschrift für Erziehungswissenschaft 14* (Suppl. 2), 67–86.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Wu, M., Adams, R., & Haldane, S. (2007). ConQuest [computer software]. Melbourne: Australian Council for Educational Research.